

On understanding word order asymmetries (Comments on John A. Hawkins, "Symmetries and asymmetries: their grammar, typology and parsing")

MARTIN HASPELMATH

1. Language structure is made for language users

John Hawkins's ideas about the explanation of word order phenomena in the world's languages have always struck me as highly persuasive and full of insights, and the target article is no exception. The success of Hawkins's work is particularly amazing in view of the fact that he has few fellow functionalists to discuss his theories with. Functionalists tend to be more comfortable with aspects of grammatical structure that have semantic content, and with grammatical asymmetries that can be attributed to the workings of economy or iconicity principles. Constituent order, which used to be hotly debated in functional-typological circles in the 1970s and early 1980s, is now usually left to generative grammarians. And indeed the array of descriptive devices that the Chomskyan framework makes available can be easily applied in discussions of word order facts of particular languages, but to the extent that Hawkins's generalizations are captured at all in generative work, this mostly happens through stipulations (e.g. the fact that *wh*-phrases cannot be moved to the right is usually attributed to a principle of UG, but in the absence of independent evidence that a language with rightward *wh*-movement is unlearnable, this amounts to a stipulation). The basic idea that language structure is made for its users, i.e. that structure is functionally adapted to language use, leads Hawkins to formulate explanations that provide far deeper insights.

Over the last decade, many linguists have become attracted to the new formalism of Optimality Theory, probably because OT allows the use of widely accepted principles (such as NoCODA, STRESS-TO-WEIGHT in phonology, STAY and DROP TOPIC in syntax) in the formalization itself, and it has often been noticed that to the extent that these principles can be thought of as functional motivations, OT is similar to the

competing-motivations analyses of functional linguistics (e.g. Hurch 1998). For me, OT is at its best when a “factorial typology” is derived from different rankings of the available constraints and it is shown that all possible rankings yield attested language types. Hawkins presents a similar factorial typology for verb-object order and relative-clause-noun order in (31) in his §4.1, which could be formulated in OT terms in the following way, using MiD (Minimize Domains) and MaOP (Maximize On-line Processing) as OT-style constraints:

- | | | | | |
|-----|-----------|----------------|-------------|----------------------|
| (1) | VO & NRel | (e.g. English) | MiD, MaOP | (unranked) |
| | OV & RelN | (e.g. Turkish) | MiD >> MaOP | (MiD ranked higher) |
| | OV & NRel | (e.g. Persian) | MaOP >> MiD | (MaOP ranked higher) |

No possible ranking of MaOP and MiD gives rise to the pattern VO & RelN, so it is predicted (correctly) that this is (almost) unattested. The main problem with such explanations in the OT framework is that they only work if it can be guaranteed that the opposite constraints do not exist (i.e. “Maximize Domains” and “Minimize On-line Processing”). Since standard OT does not have any constraints on constraints, the explanation is incomplete (cf. Haspelmath 1999). Hawkins, by contrast, motivates his principles in psycholinguistic terms, so the explanatory principles are themselves constrained.

Still, even OT practitioners who look favorably at functionalist work and see their analyses as complementary to it (e.g. Aissen & Bresnan 2002) will hesitate to adopt principles such as MiD and MaOP into their constraint set. Not only is the evaluation of MiD and MaOP a rather intricate business (see, e.g., Hawkins’s (19)–(20), (26), (30)), but perhaps more crucially, it involves quantification. But quantification of linguistic units is generally regarded as outside of the domain of competence by generative grammarians: “Grammars don’t count.” It seems quite likely that this is indeed a correct principle, and I regard it as one of the best candidates for an innate principle of UG. However, Hawkins’s principles MiD and MaOP are not competence principles, but performance principles, and hence there is no contradiction here as long as we recognize that “parsers do count”. The explanation thus crucially involves performance preferences and their conventionalization in diachronic change (i.e. diachronic adaptation). In Newmeyer’s (2002) terms, Hawkins’s

functional explanations are “holistic”, in contrast to “atomistic” functional explanations as assumed in functional OT.

Of course, Hawkins’s approach is not totally new, and in Hawkins (1994: 118) he himself cites older work by Otto Behaghel as an antecedent. Also Otto Jespersen’s work of about a century ago contains some of the basic insights. And Hawkins’s explanation of the association of rich agreement and verb-initial order is found already in Nichols (1992: 108). However, he must get the credit for formulating the general principles and exploring them in detail, linking them to psycholinguistic research and testing their predictions for synchronic variation and for typological patterns.

In what follows, I will first look at some possible extensions of the principle MaOP (§2), before I go on to ask a few critical questions: What are the limits of “dependency”-based explanation? (§3), and what are the respective roles of hearers vs. speakers? (§4) Further foundational issues are raised in §5, and finally (in §6) I ask why the processing theory of word order is still so widely resisted.

2. Further possible extensions

For the interpretation of an adjective, it makes a great difference whether it occurs in a comparative construction or not. In the sentence *Fido is big*, *big* is understood in an “absolute” sense as ‘big for a dog’, but in *Fido is bigger than X*, only the dimension of bigness is predicated, and if X is, say, a rabbit, then this means that Fido may actually be quite small for a dog. Thus, when the adjective shows no comparative-degree marking and the standard (i.e. the ‘than’-phrase) follows the adjective (as in Hebrew, cf. (2)), the special “relative” meaning of the adjective remains unassigned until the standard is parsed.

(2) Hebrew

Fido gadol min Parnavon.

Fido big than rabbit

‘Fido is bigger than a rabbit.’

(3) Kanuri (Cyffer 1991: 86)

Musa Ali-ro kura won.

Musa Ali-than big is

‘Musa is bigger than Ali.’

If, however, the standard precedes the adjective (as in Kanuri, example (3)), then the special “relative” meaning is available immediately. These considerations lead us to expect the following morphosyntactic asymmetry: In languages with adjective-standard order, special marking of the adjective (like the *-er* suffix in English) is preferred because it is needed to signal the “relative meaning” right away, whereas in languages with standard-adjective order, no special marking of the adjective is required. In other words, languages like English and Kanuri should predominate over languages like Hebrew with its relatively inefficient structure. I know of no world-wide studies of comparatives that take degree marking on the adjective into account (Stassen 1985 concentrates on other properties of comparatives), but at least in Eurasia it seems that the head-initial languages of Europe and the Middle East have degree marking, whereas the head-final languages further to the east lack degree marking (Haspelmath 2001: 1501–2).

The situation with equative constructions in Europe is similar: In languages with adjective-standard order such as Icelandic (cf. 4), we find a special equative-degree marker (*jafn* in Icelandic), whereas languages with standard-adjective order such as Kalmyk (cf. 5) lack such a marker (Haspelmath with Buchholz 1998: 294–97).

(4) Icelandic

Systir mín er jafn stór og ég.

sister my is as tall as I

‘My sister is as tall as I.’

(5) Kalmyk

Endr öskldür s&ingkiitn.

today yesterday as cold

‘Today it is as cold as yesterday.’

The evidence here is not conclusive because we have no world-wide data, but it seems quite likely that these partial patterns are not accidental. I came up with this explanation quite independently of Hawkins’s work, so the fact that it can be so easily subsumed under his MaOP principle is encouraging.

The other asymmetry that I want to mention here is much better known and well-documented: It is the cross-linguistic tendency to put the negative word early in the clause. The observation goes back at least

to Jespersen (1917: 5)¹, and Dryer (1988: 102) provides the cross-linguistic data: In his sample of 325 languages, 227 languages (or 70%) place the negative before the verb. Dryer formulates the motivation for this tendency very clearly:

"It seems plausible that this tendency is motivated by the unusual communicative properties of negative morphemes. Negative morphemes carry a large communicative load in the sense that they carry an important part of the message. If a hearer fails to hear the negative morpheme in a sentence, they will have fundamentally misunderstood the sentence...it also makes sense that negative morphemes will serve their purpose more effectively if they are not postponed until after the verb. delayin them increases the risk of misunderstanding, creating a kind of 'semantic garden path', since the apparent meaning of a sentence up to but not including the negative will be the opposite of the intended meaning."

I find this explanation quite plausible (see also my application of the same principle in accounting for an ordering asymmetry concerning indefinite pronouns, Haspelmath 1997: 205–221), and it is closely related to the explanations that Hawkins offers. However, there are also some differences. Let us take a typical example of a 'semantic garden path' sentence involving a postposed negative, the (poetic) German sentence *Das Leben ist der Güter höchstes nicht* ('Life is not the greatest of goods') cited by Jespersen (1917: 10). Do we want to say that the property "negation" is unassigned till the last word (*nicht*) of this sentence? But unassigned to what? In the target article, Hawkins mostly talks about assignment of syntactic and semantic properties to a particular element that is unspecified for the properties. In the case of negative markers, it is not so clear that we want to say that negation is a property of the verb for which it is unspecified. In some intuitive sense, negation concerns the whole clause, or at least the whole VP, and it is also not evident that negation is a "property" of something else rather than a simple meaning constituent. So if we want to apply the MaOP principle here, we need to clarify a few more concepts. Moreover, Jespersen's and Dryer's claim that negation is particularly important, more so than many other elements of the clause, does not yet find a natural place in Hawkins's conceptual apparatus. The MaOP principle says that on-line property assignments should be maximized, but not that this may be more important for some

¹ "There is a natural tendency, also for the sake of clearness, to place the negative first, or at any rate as soon as possible, very often before the particular word to be negated (generally the verb)."

properties (or constituents) than others. This is not meant as a criticism of Hawkins's theory, because he does not claim that his principles are exhaustive, but rather as a reminder that it only constitutes the beginning of our understanding of these phenomena.

3. What are the limits of "dependency"-based explanation?

Hawkins's notion of "dependency" of unspecified properties implies that ordering of two elements A and B is expected to be asymmetric unless the two elements are mutually dependent and the strength of the dependency is the same in both directions. Thus, the ordering of adjective and noun is expected to be symmetrical because the dependency is mutual: Adjectives depend on nouns for ambiguity resolution (cf. *hot water, hot dish, hot topic*), and nouns depend on adjectives for referential reduction. Hawkins argues for mutual dependency in the case of noun and relative clause, verb and object argument, verb and PP, and adposition and complement. Conspicuously absent from this list is noun and genitive, and here I wonder how they can be argued to exhibit mutual dependence. Clearly, possessed nouns are referentially reduced by their possessors, just as they are referentially reduced by modifying adjectives. But possessors generally do not depend on their possessed nouns for ambiguity resolution, and semantically possessors function in many ways like topics (see, e.g., Langacker 1993). We would therefore expect genitive-noun order to be asymmetric, contrary to what we actually observe.

There are further predictions made by Hawkins's theory that he does not mention (perhaps just for lack of space). For instance, given that we find morphosyntactic asymmetries of the kind discussed in Hawkins's §7, we would expect to find them in other cases as well. Consider again the order of adjectives and nouns: Languages with NA order do not have the problem of ambiguity resolution, because the noun has already been processed at the time when the adjective is encountered. However, AN languages do have this forward-looking dependency, and one might expect them to do something about it (just like verb-initial languages do something to avoid forward-looking dependencies on their arguments: they display rich agreement). For instance, an AN language could reduce the amount of ambiguity found in its adjectives, and we might expect to find relatively more adjective ambiguity in NA languages. This prediction is not

so easy to test, but so far nobody seems to have suggested that Romance (predominantly NA) languages have more adjective ambiguity than Germanic (exclusively AN) languages, or that Polish (NA) has more than Russian (AN).

Similarly, when comparing OV and VO languages, we would expect to find that OV languages (where processing decisions arguments can be made immediately) allow more ambitransitive verbs such as English *break* (tr./intr.) with “valence ambiguity”, because the verb’s valence is clear from its arguments. By the same token, OV languages should allow more polysemy of verbs. I do not know whether these predictions are correct, but so far nobody has presented evidence to show that the ordering of object and verb correlates with such lexical properties of verbs.

The possibilities of “dependency”-based explanation are intriguing, but it seems that at the moment we should regard these claims with some caution.

4. Hearers vs. speakers

How does parsing efficiency lead to performance preferences? Is it because speaking is made easier by parsing efficiency or because speakers want to help hearers in decoding their messages? Hawkins says nothing about this: In his article about structural asymmetries, the pragmatic asymmetry of speakers and hearers is left out of the picture. At least in the case of garden path phenomena, it would seem to be clear that the “parsing difficulty” primarily affects hearers, because garden path sentences present no special problem for speakers. So at first I interpreted Hawkins as talking about structural decisions that affect hearers. However, in his discussion of verb-object ordering, Hawkins also mentions case-assignment as a type of dependency of arguments on the verb. He must be referring to cases like German *das Kind_{ACC} unterstützen* ‘support the child’, versus *dem Kind_{DAT} helfen* ‘help the child’. This is indeed a property of the argument that depends on the verb (the verb is sometimes said to *govern* the argument’s case), but crucially it is not a covert property that the hearer has to reconstruct from the signal (unlike syntactic nodes, semantic roles, and other covert pieces of linguistic information). If the verb precedes the argument that it governs (*hilft dem Kind_{DAT}*), this presumably helps the speaker because the decision about the argument’s case only has to be made after the verb is uttered. However, for the hearer the order of argument

and verb is irrelevant because identifying the argument's case is easy: It is overt and hence never remains unassigned.

The phenomena discussed by Hawkins do not provide compelling evidence that structural properties may also be motivated by speakers' needs. He mentions case-assignment only in connection with semantic-role assignment, so it is possible that only the latter is relevant and case-assignment can simply be disregarded. However, I can think of one type of asymmetry that seems to require that we take speakers' needs seriously, too. Although I know of no systematic surveys, there is apparently a tendency for modifier-noun agreement to occur more systematically when the modifier follows the noun than when it precedes it. This is expressed in the following generalization (Moravcsik 1995: 472):

- (6) In all languages that mark a constituent preposed to the noun for inflectional categories such as number, gender, and case, constituents postposed to the noun will also carry such markings.

There are quite a few cases where we find an alternation of the type (7a) within a single language (i.e. the modifier only exhibits agreement when it is prenominal), whereas alternations of the type (7b) seem to be unattested or very rare.

- (7) a. $[\text{Mod } N_x]_{NP} \sim [N_x \text{ Mod}_{AGR-X}]_{NP}$
 b. $*[N_x \text{ Mod}]_{NP} \sim [\text{Mod}_{AGR-X} N_x]_{NP}$

It is hard to see how this asymmetry should be motivated by the hearer's needs. If the agreement-bearing modifier preceded the noun (as in the pattern $[\text{Mod}_{AGR-X} N_x]_{NP}$), then that would provide some basic information about the noun early on, but this cannot be better for the hearer than putting the noun itself first (as in $[N_x \text{ Mod}_{AGR-X}]_{NP}$). By contrast, the pattern in (7a) is clearly good for the speaker: If the modifier precedes, it does not have to be supplied with forward-looking agreement inflection, whereas such inflection is relatively easy to add if the noun precedes the modifier and has therefore been uttered by the time the modifier is produced.

But now note that if it is true that speakers' needs also influence language structure, it is no longer possible to explain Hawkins's generalization in (50), because rich verb agreement in verb-initial languages, while beneficial for the hearer, is bad for the speakers. These two motivations should cancel each other out. This is thus a puzzle that presents an interesting challenge for future research.

5. Some foundational issues

One foundational question that arises immediately is: How deep does the explanation go? In particular, can we explain the two parsing preferences, Minimize Domains and Maximize On-line Processing? Are these arbitrary properties of the human parser that could just as well be otherwise? Or can they somehow be reduced to even more general principles such as economy (or ‘minimize production effort’) and clarity (or ‘maximize ease of decoding’)? Since the two parsing preferences are to some extent supported also by evidence from outside linguistics (i.e. psychology), one could say that the linguist need not worry about these deeper questions because they lead us outside of linguistics. But to the extent that the parsing preferences can themselves be explained, I would say that it would make their explanatory role for linguistics even more convincing.

Another topic that is largely left aside by Hawkins is the nature of the conventionalization of performance preferences. He sometimes uses metaphorical terminology when talking about this: “grammars have *responded* to the efficiency differences” (p. 111); “The conventions, as I see them, are largely ‘frozen’ or ‘fixed’ performance preferences” (p. 146) (*italics added*). But how exactly does the conventionalization (the “responding”, the “freezing”, the “fixing”) work? Hawkins says nothing about this and instead refers to Haspelmath (1999) and Kirby (1999), who ask the same questions but come up with rather different answers. In Kirby’s view, structures with low parsing efficiency tend to disappear from a language because they are not easily acquired by children. Haspelmath, by contrast, emphasizes the role of the speaker: Speakers prefer structures with high parsing efficiency (to aid the hearer in decoding), and disfavored structures disappear because they become too rare. However, Croft (1999) objects to this view that there is no evidence for “functional selection” in language change (and that all selection is instead “social selection”), claiming that the observed result of functionally adapted structures is due to the tendency for innovations to be adaptive. The problem with Croft’s approach is that many innovations consist in changes of relative frequency and in the loss of already disfavored variants, and it is difficult to see how such changes could spread by exclusively social selection. There is no space here to discuss these issues in detail, but they must be taken seriously if we are aiming for a complete explanation of linguistic patterns. It may well be that the process of conventionalization imposes additional constraints on the kinds of grammars that result from it, and if we simply

ignore this process, we run the risk of putting too many constraints in the parser.

6. Why is the processing theory of word order still resisted?

In my view, Hawkins's theory of word order has not been sufficiently appreciated by linguists, who instead often prefer approaches that attempt to attribute word order phenomena to Universal Grammar in a highly speculative and haphazard way. If I am right that Hawkins's approach is by far the most fruitful one of those currently on the market, why hasn't it swept the field? The answer to this question has many facets, many of them having to do with accidents of the sociology of linguistics that need not be discussed here. Here I would just like to highlight one aspect that I find rarely mentioned by linguists.

The best theories are those that account for the widest range of data, i.e. not just for a few languages. Since his early work in the late seventies, Hawkins has based his theoretical research on the results of empirical word order typology, and while a lot of new facts have been uncovered since Greenberg's early work, his main generalizations still stand unchallenged, and theories based on these studies are on fairly safe ground. However, most linguists are not typologists but try to say something new about individual languages that they happen to be specialists of. If one's task is to write a research paper about word order in language X, then adopting a particular framework such as Chomsky's latest framework is easier than trying to learn something significant about the language from a typological theory. In many cases, the formal framework was developed using some other language(s) (often English), so that a framework-based approach often yields problems right away that can then be tackled by the researcher. The paper will then usually propose a slight adaptation of the framework, and as a result both an elegant description of the particular facts ("an analysis") and a small contribution to the general theory is offered.

This has proved an efficient method to generate research papers and dissertations, and no doubt many important details about languages have been discovered in this way. But it has not led to a general explanatory theory of word order that would be able to compete with Hawkins's. And it still remains the case that the facts of individual languages are to a large extent accidental and will always resist explanation. True

explanations can only be given for universals, and these require a typological approach.

So my explanation for the continued popularity of formal frameworks is that they allow specialists of individual languages at least the illusion of being little theorists of their own, and unfortunately for linguistics, theoretical linguistics seems to be widely regarded as more prestigious than descriptive linguistics (this is in contrast to physics, where experimental physics seems to be more prestigious than theoretical physics, perhaps because it usually involves working with higher budgets).

However, the performance predictions of Hawkins's theory lend themselves far more easily to verification through in-depth studies of individual languages than the competence predictions, and it seems safe to predict that the ever increasing technological possibilities for corpus research will lead to a greater popularity of theories that have something interesting to say not only about grammaticality, but also about discourse frequencies. So in the long run, I am optimistic about the impact of ideas like Hawkins's on linguists in the field.

Max-Planck-Institut für evolutionäre Anthropologie, Leipzig
haspelmath@eva.mpg.de

References

- Aissen, Judith & Bresnan, Joan (2002). Optimality and functionality: Objections and refutations. *Natural Language and Linguistic Theory* 21
- Croft, William (1999). Adaptation, optimality and diachrony (commentary on Haspelmath (1999). *Zeitschrift für Sprachwissenschaft* 18.2, 206–208
- Cyffer, Norbert (1991). *We learn Kanuri*. Köln: Rüdiger Köppe Verlag
- Dryer, Matthew (1988). Universals of negative position. In: Michael, H., Edith, M. & Jessica, W. (eds.). *Studies in syntactic typology*. Amsterdam: Benjamins, 93–124
- Haspelmath, Martin (1997). *Indefinite pronouns*. (Oxford Studies in Typology and Linguistic Theory) Oxford: Oxford University Press
- Haspelmath, Martin (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18.2, 180–205
- Haspelmath, Martin (2001). The European linguistic area: Standard Average European. In: Martin, H., Ekkehard, K., Wulf, O. & Wolfgang, R. (eds.). *Language typology and language universals*. (Handbücher zur Sprach- und Kommunikationswissenschaft) Berlin: de Gruyter, 1492–1510
- Haspelmath, Martin with Oda Buchholz (1998). Equative and similitive constructions in the languages of Europe. In: Johan, van der A. (ed.). *Adverbial constructions in the languages of Europe*. (Empirical Approaches to Language Typology/EUROTYP, 20-3.) Berlin: Mouton de Gruyter, 277–334

- Hawkins, John A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press
- Hurch, Bernhard (1998). Optimalität und Natürlichkeit. *ZAS Papers in Linguistics* 13, 115–139 (Berlin: Zentrum für Allgemeine Sprachwissenschaft)
- Jespersen, Otto (1917). *Negation in English and other languages*. Copenhagen: Bianco Luno
- Kirby, Simon (1999). *Function, selection, and innateness: The emergence of language universals*. Oxford: Oxford University Press
- Langacker, Ronald W. (1993). Reference-point constructions. *Cognitive Linguistics* 4.1, 1–38
- Moravcsik, Edith A. (1995). Summing up Suffixaufnahme. In: Frans, P. (ed.). *Double case: agreement by Suffixaufnahme*. New York: Oxford University Press, 451–484
- Newmeyer, Frederick J. (2002). Where is functional explanation? *Papers from the Thirty-Seventh Meeting of the Chicago Linguistic Society*, ed. by Andronis, M., Ball, C., Elston, H. & Neuvel, S. Chicago: Chicago Linguistic Society
- Nichols, Johanna (1992). *Linguistic diversity in space and time*. Chicago: The University of Chicago Press
- Stassen, Leon (1985). *Comparison and universal grammar*. Oxford: Blackwell